*Reviewer #1: I thank the authors for their simple but useful generalization of weighted UniFrac distances from relative to absolute abundances. The key concepts are explained well and the properties and utility of the metrics are demonstrated and compared to others through toy examples as well as a real-world dataset. In general, it would be interesting to see comparison of metrics for more than one example, but I guess there are strict limitations for a brief communication.*

We thank the reviewer for their interest and encouragement. We have added substantially to the paper, including three additional data sets within our analysis, and an extended discussion (moving outside the constraints of the brief communication).

*My only major comment is as follows:  
  
To me, it seems trivial to replace relative abundance by absolute counts. Is this the first time this has been suggested or tried? Why isn't this already commonly used, given that suitable absolute abundance data is available?*

This is a useful comment, and reflects what was to us as a surprising gap in the literature. We added additional information (line 70) to emphasize that these metric has (to our knowledge) not been used, and that while its derivation and application are simple, the resulting interpretation is nontrivial:

“As can be seen, this replacement is mathematically trivial, but we were unable to find any examples in the literature where absolute abundances are discussed in relation to Unifrac distances, either conceptually or in application. Absolute abundances add another axis along which can vary; besides differences in community composition and phylogenetic similarity, also considers absolute differences. As such, it is a nontrivial task to consider how behaves in comparison to other metrics of using -diversity.”  
  
*In addition to this, I have some minor comments that should be addressed before publication:  
  
99: It's informative to see how UA relates to the other metrics for specific examples, but I find it hard to draw the conclusion that it "integrates ecological realism" just from the numbers that are provided. Could you substantiate this claim more? I am also curious about how interpretable UA is compared to other metrics?*

We agree slightly more nuance is needed within that claim. We’ve updated the text to reflect that UA is able to integrate multiple axes of important ecological variation, but that, as a result, its interpretation is more complex. This leads well to the additional analyses we present using our new datasets, e.g.:

“These scenarios demonstrate that can integrate changes along multiple, ecologically relevant axes (including differences in composition, phylogenetic similarity, and microbial load). However, because a given distance is the result of multiple drivers of variation between communities, its interpretation likely requires additional data analyses – are differences being driven primarily by changes in absolute abundance, or composition?

To illustrate the sensitivity of to variation in composition and absolute abundance, we re-analysed four previously published datasets from diverse microbial systems.”

*Fig. 1B: Why aren't all the colored stars indicated in all the distributions in B?*

The authors had originally added stars to the histograms sparingly, to draw attention to specific comparisons. However, stars have been added to all panels now.

*Fig. 1C: Maybe I'm misunderstanding something, but aren't the first two trees in C (gold star) exactly the same, including the absolute abundances? In that case, shouldn't distances and dissimilarities be zero? Is there an error in the labels of one of the trees?*

This mistake has been corrected.

*108: "UA yields greater or smaller dissimilarity than other metrics". UA and UR are the same in the first example.*

We’ve clarified this sentence, so that it’s not declarative of UA’s relation to all other metrics, but reflects that these scenarios were chosen to illustrate specific cases where UA differs from at least one other metric in a conceptually important way. The new sentence reads:

“Examples of illustrative comparisons which reflect how absolute abundances and phylogenetic similarity can drive differences across dissimilarity metrics.”

*142: "In this dataset, we recommend an intermediate    of 0.5". This seems a bit arbitrary, and isn't the horseshoe still discernible for alpha = 0.5? Why do you recommend this value and how do you recommend choosing alpha in general?*

We’ve updated our analyses to include a finer range of alphas (in steps of 0.1, rather than in 0.5), and included a new guidance for choosing an alpha, with general recommendations to 1) Explore multiple levels of alpha; 2) Consider choosing alpha *a priori* to modulate the impact of cell abundance on your diversity estimation; and, 3) Choose an alpha in an intermediate range (0.25-0.75) but without a specific recommendation of 0.5.

*158: "For example, the temporal development of the infant microbiome". It would have been interesting to see the different metrics applied to data for this example.*Unfortunately we were unable to access the necessary data to reanalyze that study. However, (as discussed elsewhere) we’ve incorporated analyses from three other datasets across a broad range of environments, richness, and evenness. *Reviewer #2*

*Major Comments:  
  
As indicated by your own words on lines 81-83, "These comparisons emphasize that incorporating phylogeny and absolute abundance reshapes distance estimates in nontrivial ways.", there are novel properties associated with the newly proposed Absolute Unifrac and Generalized Absolute Unifrac metrics. These need to be understood through more than a toy four-ASV phylogeny and one freshwater 16S data set. It would be ideal to at least test:*

*1) data derived from multiple environments (e.g. soils, human gut) that represent a range of species diversity and abundances,*

We’ve incorporated three additional datasets ranging widely in richness and abundances, which we feel has greatly improved the manuscript.

*2) data derived from other marker genes (e.g. 18S, cytochrome oxidase, etc.),*

We agree that more analyses would be useful to the field. That said, it was already difficult to find studies with publicly available sequencing data, absolute abundance measurements, and high-quality metadata that were acceptable for reanalysis. Studies incorporating absolute abundance are few and far between; studies with well-reported data are also unfortunately not as common as they should be. While one dataset we accessed (Zhang et al., peanut rhizosphere) did include fungal ITS sequences, we felt it wasn’t additive to compare these results to the bacterial 16S results, without other ITS or 18S studies with which to compare it.

*3) how the metric's validity is impacted by precision and error at the qPCR or flow-cytometry steps that could propagate into Absolute Unifrac distances,*

We agree this is an important point. While a mathematical analysis of error propagation in UA vs. BCA is outside our expertise, we include simple simulations wherein error was added to absolute abundance measurements and GUA/BCA were re-calculated (Fig. 4A/B and corresponding results text, lines XXX), demonstrating the general resilience of these methods to quantification error.

*4) how this metric is impacted by rarefaction decisions.  
  
On the last point re: rarefaction but also related to other points, the authors note on line 171 that "We also do not address…how sequencing depth influences richness estimates or whether rarefaction should be applied before calculating GUA.", but given that this is the paper introducing this new metric, and given that the metric directly relies on the abundance values which are the target of rarefaction efforts, it seems reasonable to expect guidance for readers on these steps. Likely, more testing is required to assess the impact of rarefaction on various data sets employing GUA, but this key effort would lead to the creation of the guide readers require to apply this new computation tool.*

We agree that the impact of rarefaction is important, and that a full analysis of rarefaction with this metric is outside the scope of this article. However, in the spirit of readers using this manuscript as a guide for their own analyses, we include a conceptual overview of our rarefaction approach (Fig. S4) alongside corresponding results text (lines XXX), and available code for its implementation (Github link) and encourage future studies on this subject.

*If such a guide could be produced to highlight the use cases for this new metric, it should be published. Ideally, everyone understands the conditions under which this metric outperforms other abundance-aware distances, of which there are many.*

This comment was thought-provoking and shifted our presentation of these results. No single metric necessarily “out-performs” another, as each metric measures different, but valid, axes of variation between microbial communities. We added discussion text (final paragraph) highlighting that the choice of metric–and even more critically, its interpretation–should be driven by what hypotheses a researcher is seeking to test, and whether they relate to compositional similarity, phylogenetic similarity, or differences in absolute abundance. As a whole, the manuscript demonstrates that UA effectively incorporates all three of these components of microbial communities, and provides examples of how researchers can interpret its results. *Minor Comments:  
  
The simulation data used for the phylogeny is far too simple and does not demonstrate ecological realism. Related to the comments about testing more data sets, the in-silico data can certainly be more complex and robust as a test data set.*

The purpose of Fig. 1 is illustrative, rather than analytical.The four-ASV community was chosen because it is the simplest possible demonstration of how these metrics can respond to changes in composition, abundance, and phylogenetic similarity. We feel adding complexity to this figure would work against its purpose, and think that the inclusion of new, real-world datasets better address these concerns, especially as they recapitulate many of the patterns that we demonstrate even with the exceptionally simple 4-ASV community. *The data and code available to produce the manuscript are all seemingly shared in a public Github. Kudos to the authors for this important step in providing a reproducible analysis.*

Thank you! We’ve maintained this degree of reproducibility throughout all reanalyses as well. *If possible, for when data exists for a range of diverse data sets, it would be valuable to add any statements about computational efficiency relative to existing metrics (e.g. Bray-Curtis, Unifrac, Weighted Unifrac, etc.).*

We’ve included a new figure (Fig. 4 and Fig. S5) which demonstrate how GUnifrac is specifically slower than Bray-Curtis (as performed via the vegan package) and Fast Unifrac (as calculated within the phyloseq package). We’ve also included details in the discussion point to the >1000x slow-down represented by GUnifrac, but make some recommendations for how it could be sped up across iterations of rarefaction. *Figure S2 could indicate the sample labels, or at least a key for the labels, to aid in interpretability.*

As we’ve expanded to multiple datasets, we no longer include what was previously Fig. S2 and S3. The ordination based on UR (old Fig. S3) is now included as a sub-panel in Fig. S3, which demonstrates ordinations using UR and UA across all four datasets.